COD, PCOD, TCOD, MPOD …

Open structure and property databases

Daniel Chateigner[1,*], Saulius Grazulis[2], Olivier Pérez[1], Giancarlo Pepponi[3], Luca Lutterotti[1,4]

1.CRISMAT-ENSICAEN-CNRS, IUT Caen,Normandie Universités, Caen, France
2.Institut of Biotechnology, Vilnius, Lithuania
3.Fundazione Bruno Kessler, Trento, Italy
4.Dept. of Materials Engineering, University of Trento, Italy
** daniel.chateigner@ensicaen.fr*

Résumé abrégé : Open databases are developping for easy and free phase identification, phase quantifications from powder papers, predicted structures for materials engineering and prospects, and tensor properties.
Mots-clés : Open database, quantitative phase analysis, tensor properties

The COD project (abbreviated from the "Crystallography Open Database", http://www.crystallography.net/) aims at collecting in a single open access database all organic, inorganic and metal organic structures [1]. The database was founded by Armel Le Bail, Lachlan Cranswick, Michael Berndt, Luca Lutterotti, Daniel Chateigner and Robert M. Downs in February 2003 as a response to Michael Berndt's letter published in the Structure Determination by Powder Diffractometry (SDPD) mailing list [2]. Since December 2007 the main database server is maintained and new software is developed in the Vilnius University Institute of Biotechnology by Saulius Gražulis and Andrius Merkys, and has now over 220 thousand records describing structures published in IUCr journals, other major crystallographic and chemical peer-reviewed journals [3], and individual contributors. Structure file harvesting is now routinely operated and automated for a growing number of publishers' journals.

The database presents itself on the Internet as a Web site (Fig. 1a) with the basic data search and download capabilities, designed by Armel Le Bail and Michael Berndt. In addition, registered users may deposit new data into the database, either form the previous publications or as personal communications, using the deposition web site designed in the VU Institute of Biotechnology by Saulius Gražulis, Justas Butkus and Andrius Merkys. The deposition software performs rigorous checks of syntax and semantics, thus ensuring high quality of records deposited in the COD.

The COD web site allows searching on COD numerical identifier, unit cell parameters, chemical composition, and bibliographic data. Substructure searches using SMILES and SMARTS strings have been implemented. The free software package Openbabel [4,5] is used for both the CIF-to-SMILES transformation and the actual search.

The retrieved COD records can be viewed on-line (Fig. 1b) or downloaded for further processing. For massive data mining, COD permits downloads and updates of the whole database using Subversion, Rsync or HTTP protocols. The ease of access to COD data has spurred the use of this resource for software testing [6], teaching [7], and research [8].

For the powder diffractometry community, COD is interesting not only as an archive of structures solved by powder diffraction methods, but also as a possibility to use it in existing search-match procedure. A recent development of an open Full-Pattern Search-Match internet tool was also launched. It allows phase quantification from x-rays, neutron and electron powder patterns, with high or medium resolution instruments, provided the structures are already in COD. This tools is particularly suited for nanocrystalline powders in which severe line broadening appears precluding phase identification from only peak maxima (cod.iutcaen.unicaen.fr [9]). COD-derived databases are offered for software produced by many diffractometer vendors. In addition to COD, search match can be done against its sister database, PCOD, that contains structures predicted by the GRINSP program [10]. Soon will be launched the new TCOD database, which will collect structures optimized from COD using first-principles calculations like DFT. The open nature of the COD permitted numerous mirrors around the globe [11-14] and specifically tailored COD database variants

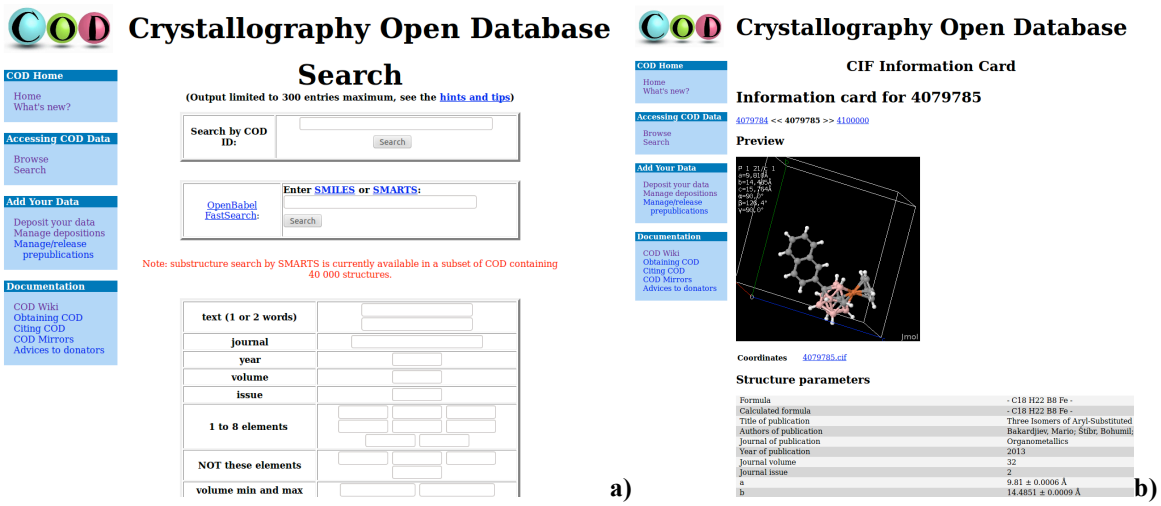

Fig. 1 a) Web site and search interface of the Crystallography Open Database (COD) permits searches of crystallographic data by a range of parameters and unrestricted retrieval of the found data. b) data can be viewed on-line in the Jmol [9,10] applet or downloaded for further processing, either record-wise or in bulk.

Meanwhile, the Material Properties Open Database got created in 2010 [15], which collects tensor properties. This latter sister is on a demanding growth and is actually skewed towards elastic properties. Optical properties, if already targeted require more investment, to come.

References

1. Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P. & Le Bail, A. (**2009**). *Crystallography Open Database -- an open-access collection of crystal structures*, Journal of Applied Crystallography 42 : 726-729.
2. Berndt, M. (**2003**). *Open crystallographic database - a role for whom?*, http://tech.groups.yahoo.com/group/sdpd/message/1016 (retrieved 2013.01.31).
3. Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T. & Le Bail, A. (**2012**). *Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration*, Nucleic Acids Research 40 : D420-D427.
4. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T. & Hutchison, G. R. (**2011**). *Open Babel: An open chemical toolbox*, Journal of Cheminformatics 3 : 33.
5. Hutchinson, G. (**2007**). *Open Babel: The Open Source Chemistry Toolbox*, http://openbabel.org/ (retrieved 2013-04-19).
6. Grosse-Kunstleve, R. & Gildea, R. (**2011**). *Computational Crystallography Initiative: COD stats*, http://cci.lbl.gov/cod_stats/ (retrieved 2013.01.31).
7. Moeck, P. (**2004**). *EDU-COD: Educational Subset of COD*, http://nanocrystallography.research.pdx.edu/search/edu/ (retrieved 2013.01.31).
8. First, E. L. & Floudas, C. A. (**2013**). *MOFomics: Computational pore characterization of metal–organic frameworks*, Microporous and Mesoporous Materials 165 : 32-39.
9 L. Lutterotti, H. Pilliere, C. Fontugne, P. Boullay, D. Chateigner (**2013**), in preparation
10. Le Bail, A. (**2005**). *Inorganic structure prediction with it GRINSP*, Journal of Applied Crystallography 38 : 389-395.
11. Quirós-Olozábal, M. (**2006**). *COD Mirror of Granada University*, http://qiserver.ugr.es/cod/ (retrieved 2013.01.31).
12. Moeck, P. (**2007**). *Crystallography Open Database Mirror*, http://nanocrystallography.research.pdx.edu/search/codmirror/ (retrieved 2013.01.31).
13. Gražulis, S. (**2007**). *COD Mirror in Vilnius*, http://cod.ibt.lt/ (retrieved 2013.01.31).
14. Chateigner, D. (**2010**). *Crystallography Open Database Mirror at ENSICAEN*, http://cod.ensicaen.fr/ (retrieved 2013.01.31).
15 G. Pepponi, S. Grazulis, D. Chateigner (**2012**). MPOD: a Material Property Open Database linked to structural information. *Nuclear Instruments and Methods in Physics Research B* **284**, 2012, 10-14. http://www.materialproperties.org/